

Copyright © 1986  
Edward Arnold (Publishers) Ltd, 41 Bedford Square, London WC1B 3DQ

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic photocopying, recording or otherwise, without the prior permission of Edward Arnold (Publishers) Ltd.

ISSN 0265-5322

Typeset by Colset Private Ltd, Singapore  
Printed in Great Britain by Paradigm Print, Gateshead.

# Language Testing

## Contents

### Articles

Academic demands related to listening skills

*Donald E. Powers*

1

The Syntactic Density Score revisited: which of its components matter in the oral language of 9-15 year olds?

*Rochelle B. Simms and Donald J. Richgels*

39

Models of second language competence: a structural equation approach

*F. Sang, B. Schmitz, H. J. Volmer, J. Baumert and P. M. Roeder*

54

Are English language aptitude tests biased towards culturally different minority groups? Some Israeli findings

*Moshe Zeidner*

80

Student affective reactions to oral language tests

*Mary Lee Scott*

99

### Book reviews

Mercado, C. I., Romero de Ortiz, M. and Vazquez, J. A.

*Conceptual framework for a comprehensive language assessment program* (reviewer: Andrew D. Cohen)

119

Kirsch, J. editor, *Tests and testing for bilingual children: a bibliography of literature*

Vasquez, J. A., Gonzales, S. E. and Pearson, M. E. *Testing and ethnic minority students: an annotated bibliography* (reviewer: Don Porter)

### Meetings

123

### Publications received

124

classroom examinations. *Journal of Abnormal and Social Psychology* 50, 93-98.

Savignon, S.J. 1972: *Communicative competence: an experiment in foreign language teaching*. Philadelphia, Pennsylvania: The Center for Curriculum Development Inc.

Schwarzer, R., van der Ploeg, H.M. and Spielberg, C.D. 1982: Test anxiety: an overview of theory and research. In Schwarzer, R., van der Ploeg, H.M. and Spielberger, C.D., editors, *Advances in test anxiety research Vol. 1* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Scott, M.L. 1980: The effect of multiple retesting on affect and test performance. Brigham Young University, unpublished MA thesis.

Scott, M.L. and Madsen, H.S. 1983: The influence of retesting on test affect. In Oller, J.W. Jr, editor, *Issues in language testing research*, Rowley, Massachusetts: Newbury House.

Shohamy, E. 1982: Affective considerations in language testing. *Modern Language Journal* 66, 13-17.

Spielberger, C.D. 1966: Theory and research on anxiety. In Spielberger, C.D., editor, *Anxiety and behavior*, New York: Academic Press.

## Appendix II Affect questionnaire

Date \_\_\_\_\_  
Level \_\_\_\_\_

Age \_\_\_\_\_  
Sex \_\_\_\_\_

This questionnaire is part of a research project being carried out by a visiting instructor from UCLA. The results of the project will be used to assess the reliability and accuracy of the oral tests for Alumni courses. For this reason it is very important that your responses be as frank as possible.

It is important to emphasize that the opinions you give are confidential and *will not interfere* in any way with your evaluation.

### Instructions

Please mark with an 'X' the description on the agreement/dis-agreement scale which best expresses your opinion. Note that in some questions you are asked to explain your response.

For example:

It is helpful to study the day before for an oral test.

/	X	/	/	/
disagree	somewhat	neutral	somewhat	agree
	disagree		agree	

1) I believe that this oral test is an accurate evaluation of my ability to speak English.

Please explain why \_\_\_\_\_

2) If I took the oral test from another teacher, I would get a different score.

Please explain why \_\_\_\_\_

3) I felt nervous before the test.

4) I felt nervous during the test.

5) I would rather take a written test than an oral test.

Please explain why \_\_\_\_\_

6) I believe I did well on the oral test.

7) If I took the same test with the same teacher, on another day, the result be would the same.

Please explain why \_\_\_\_\_

8) I believe that oral tests are necessary in English courses.

9) I believe I had an adequate opportunity to demonstrate my ability to speak English.

Please explain why \_\_\_\_\_

10) The oral test was too short.

11) I liked the oral test.

Please explain why \_\_\_\_\_

12) I understood what I was supposed to do during the oral test.

13) I thought the test corresponded to what I learned in class.

14) I would like to take the next test with the same teacher who gave it to me today.

15) I feel more comfortable when I take an oral test with another student.

16) I think the teacher who gives the midterm should also give the final exam.

17) It was very difficult to understand what the teacher said because of his/her accent.

18) I thought the oral test was too difficult.

Please explain why \_\_\_\_\_

19) I would like my classroom teacher to be present during the exam.

Please explain why \_\_\_\_\_

20) I feel more comfortable when I take an oral exam in a group, than with just one other student.

Thank you for your cooperation.

## VI References

- Alpert, R. and Haber, R.N. 1960: Anxiety in academic achievement situations. *Journal of Abnormal Psychology* 61, 207-15.
- Bruch, S. 1979: Convergent/discriminant validation of prospective teacher proficiency in oral and written production of French by means of MLA foreign language proficiency tests for teachers (TOP and TWP) and self ratings. University of Minnesota, unpublished dissertation.
- Calvin, A.D., McGuigan, F.J. and Sullivan, M.W. 1957: A further investigation between anxiety and classroom examination performance. *Journal of Education Psychology* 48, 240-44.
- Galassi, J.P., Frierson, H.T. Jr and Siegel, R.G. 1984: Cognitions, test anxiety, and test performance: a closer look. *Journal of Consulting and Clinical Psychology* 52, 319-20
- Hill, K.T. 1983: Interfering effects of test anxiety on test performance: a growing educational problem and solutions to it. *Illinois School Research and Development* 20, 8-19
- Jones, R.L., Madsen, H.S. and Brown, R.L. 1980: Evaluation of affective variables in second language testing. Unpublished paper presented at the 14th Annual TESOL Convention, San Francisco, California.
- Madsen, H.S. 1982: Determining the debilitating impact of test anxiety. *Language Learning* 32, 133-43.
- Madsen, H.S. and Murray, N. 1984: Retrospective evaluation of testing in ESL content and skills courses. Brigham Young University, unpublished manuscript.
- McKeechle, W.J., Pollie, D. and Speisman, J. 1955: Relieving anxiety in



## Appendix I Criteria for grading oral exams

## Fluency

- 1 Speech is hesitant and choppy so that conversation is *almost impossible*.
- 2 Speech is slow; sentences are left unfinished; student *frequently* rephrases.
- 3 Speech is *occasionally* hesitant with some rephrasing.
- 4 Speech is *effortless* and smooth. Student is obviously comfortable with and confident of his language.
- 5 Student is far more fluent than any other student of this level.

## Comprehension

- 1 Understands only *very simple* speech; requires constant repetition and rephrasing.
- 2 Understands carefully planned and enunciated sentences. *Often* asks for repetition.
- 3 Understands quite well; asks only *occasionally* for repetition.
- 4 Understands *natural speech* on topics covered in class.
- 5 Student understands more than what was taught in level.

## Pronunciation

- 1 Mispronunciation and inaccurate stress make understanding difficult. Student has to repeat a lot to make himself understood.
- 2 *Frequent misunderstanding* arises from student's mispronunciation of words and errors in intonation.
- 3 Mispronunciation occurs only *occasionally* and doesn't cause misunderstanding.
- 4 No obvious pronunciation problems. Student is *always intelligible*.
- 5 Student's pronunciation closely approximates that of a native speaker.

## Structure

- 1 Communication extremely difficult due to lack of control of structure taught in class. *Constant* grammatical errors.
- 2 Frequent errors showing *limited control* of structures taught in class, and causing frequent misunderstanding.
- 3 *Occasional* mistakes with familiar structures but communication is not impeded.
- 4 Very few errors. Shows *mastery* of grammar taught in level.
- 5 Student uses structures beyond what was taught in class.

## Vocabulary

- 1 Vocabulary so limited that conversation is *almost impossible*.
- 2 Word choice *not accurate* thereby making discussion of topics covered in class difficult.
- 3 *Occasional* groping for vocabulary but adequate enough for discussion of topics covered in class.
- 4 *Broad, precise* vocabulary; can cope with varied social situations.
- 5 *Extensive vocabulary* going far beyond what was taught.

when responses to the two formats were combined, there was no significant difference in the responses of students at different levels to the different test formats such as initially hypothesized (i.e. that lower level students would prefer the group format, while higher level students would favour the pair format).

The results of this study support the findings of previous studies and offer further insight into the issues which influence students' reactions to language tests. It appears that students are able to separate out their judgements of the validity of an oral test from its perceived difficulty and their anxiety in the testing situation. The factor analysis revealed two major factors governing students' responses to the affect questionnaire. One factor appeared to involve cognitive judgements about the validity of the test (including their opinion about the necessity of oral tests), and how well they liked the test. The other factor appeared to be emotive, and involved students' reports of the anxiety they felt both before and after the test.

The studies by Savignon (1972), Brutch (1979) and Shohamy (1980) also showed that test validity is an important factor in students' decisions about how well they like a test. In Savignon's study, students reacted very positively to a test even though it was difficult for them because they felt it actually tested the language skills they were trying to acquire. In the studies by Brutch and Shohamy students preferred the more communicative tests, again presumably because of their higher face validity.

In the present study, an analysis of students' explanations for their responses to selected questionnaire items found that students most frequently mentioned their emotional state in relation to the test. Many indicated that they felt nervous because of the testing situation and that this anxiety had had a detrimental effect on their test performance. Other areas of student concerns related to test content and time constraints, similar to findings reported in previous studies (Madsen, 1982; Madsen and Murray, 1984; Schwarzer *et al.*, 1982; Hill, 1983).

A few students indicated that that they would prefer that their teacher evaluate their oral performance in class over the course of the semester. Perhaps a combination of periodic classroom teacher ratings and more formal oral tests could be used to determine students' oral grades for a class, instead of basing the grades entirely on brief formal evaluations as was the policy of the language school involved in the present study.

It is apparent that there are many factors in addition to anxiety which influence reactions to language tests. Although it is important to consider students' emotional reactions when developing a test, other issues such as test validity and time constraints seem to be of

equal or even greater importance. In addition, as language programmes are most often limited in the amount of time, money and other resources that can be devoted to testing, it is important to devise tests which are cost effective in terms of their efficiency in assessing language skill, an issue not examined in the present study. Further research is necessary to determine which testing formats and/or techniques are most efficient, and at the same time have high face validity and produce the least amount of anxiety.



Table 4 - Continued

Item	Source	Mean sq	F	p
5 Prefer written test	F	.55	51	.476
	L	1.79	50	.681
	LF	.27	27	.865
6 Think performed well	F	.14	19	.667
	L	.21	12	.946
	LF	2.08	2.75	.046
7 Test is reliable	F	3.97	2.20	.141
	L	3.47	1.25	.294
	LF	2.07	1.14	.334
8 Oral tests necessary	F	.98	.98	.162
	L	1.46	1.83	.143
	LF	1.15	2.31	.079
9 Opportunity to show ability in English	F	.25	.17	.679
	L	5.16	1.92	.128
	LF	1.67	1.15	.332
10 Test too short	F	.47	.29	.588
	L	3.20	1.31	.272
	LF	1.34	.83	.477
11 Liked test	F	.14	.11	.738
	L	1.45	.75	.525
	LF	5.18	4.11	.008
12 Understood instructions	F	5.10	6.93	.010
	L	6.14	4.69	.004
	LF	1.10	1.49	.220
13 Test covered class material	F	2.64	3.12	.080
	L	5.66	5.65	.001
	LF	.82	.97	.411
14 Same teacher next time	F	6.20	6.84	.010
	L	6.63	3.76	.012
	LF	2.16	2.38	.073
15 Prefer to be tested	F	3.04	2.62	.108
	L	2.26	1.14	.337
	LF	.67	.57	.633
16 Same teacher for midterm as final	F	1.72	1.76	.187
	L	7.02	3.09	.029
	LF	1.10	1.12	.343
17 Tester accent difficult to understand	F	2.46	1.93	.168
	L	18.35	8.90	.000
	LF	.97	.76	.520
18 Test too difficult	F	2.64	2.17	.143
	L	4.38	2.19	.091
	LF	1.79	1.47	.227
19 Like classroom teacher to be present	F	.14	.21	.651
	L	2.55	.85	.468
	LF	.37	.54	.653
20 Prefer group to pair format	F	36.47	22.21	.000
	L	8.15	2.41	.069
	LF	.60	.36	.779

These factors may account for some of the differences in student responses to these items. The interaction between level and format was significant for items 1, 6 and 11, but an examination of the means failed to show a consistent pattern.

In addition to rating each questionnaire item on a five-point Likert-style scale, students were asked to explain their ratings for eight of the items (numbers 1, 2, 5, 7, 9, 11, 18 and 19). The responses for the pair and group test formats were again collapsed for this analysis.

An analysis of students' explanations for (their responses to the eight questionnaire items listed above revealed that students were most concerned about their emotional state in relation to the test. Of the total responses to the open-ended items on the affect questionnaire, a third (32 per cent) referred to this issue. A majority mentioned the negative effect of the testing situation on their emotional state and/or the detrimental effect of their nervousness on their performance.

Student concerns about test content accounted for 12 per cent of the total responses, and comments on time constraints for 11 per cent. The majority of these comments indicated that students felt there was not enough time during the test for an accurate evaluation of their oral ability in English. (In contrast, students indicated they preferred a written test because they felt it afforded them an opportunity to consider their responses more carefully. Some commented that this lack of time pressure in a written test situation resulted in a feeling of calm and greater confidence in their performance.)

Student comments about the test administrator represented only 6 per cent of their total responses. The crosstabulation discussed previously showed that this topic was the one students seemed to feel least strongly about, as indicated by the relatively high percentages of 'no opinion' responses to these items. When students did refer to the test administrator, the majority of their comments were positive. Even so, many of their responses indicated that they felt the test administrator did not influence their performance on the test.

Finally, a few students (representing 3 per cent of the total responses) indicated that they felt their oral skills could best be evaluated by the classroom teacher over the course of the semester. The remaining 36 per cent of the responses to the open-ended items were categorized as miscellaneous and didn't appear to provide additional information on concerns common to a number of students.

#### IV Conclusion

This study revealed no overall significant difference in student affective reactions to the pair vs the group test format. Although the reactions of students from different levels was statistically significant

Table 2 Principal factor solution (varimax rotated factor matrix)

Questionnaire item	Factor 1	Factor 2
1 Accurate evaluation	.636	.111
2 Different teacher = different score	-.094	.042
3 Nervous before test	-.089	.634
4 Nervous during test	-.155	.917
5 Prefer written	-.338	.299
6 Think performed well	.222	.200
7 Test is reliable	-.023	-.059
8 Oral tests necessary	.425	.024
9 Opportunity to demonstrate ability in English	.631	.148
10 Test too short	-.003	.006
11 Liked test	.478	-.060
12 Understood instructions	.133	.093
13 Test covered class material	.140	.127
14/16 Same teacher next time	.013	.096
15 Prefer to be tested w/ another student	.129	.010
17 Tester accent difficult to understand	-.083	.021
18 Test too difficult	-.154	.071
19 Like classroom teacher to be present	-.253	.010
20 Prefer group to pair format	-.010	.098

Factor 1 appears to relate to students' perception of the validity of the oral testing procedure, as well as how well they liked it. Loading positively on this factor were statements that the test seemed to be an accurate evaluation of their ability to speak English (item 1), that oral tests are necessary (item 8), that the test gave them an opportunity to demonstrate their ability to speak English (item 9), and that they liked the test (item 11). A preference for a written, rather than an oral test (item 5) loaded negatively on this factor. Factor 2 seems to relate primarily to students' feelings of anxiety, both before (item 3) and during the test (item 4).

Factor scores for the two factors were then used as dependent variables in a MANOVA to assess the significance of differences between student reaction to the pair vs group test format, between the reactions of students at different class levels, and in the interaction of these variables. Table 3 displays the results of multivariate tests of significance.

No significant difference was found among student reactions to the different test formats. Although there was a significant difference among reactions of students from different class levels when the responses to the different test formats were collapsed, an examination of the observed means failed to provide the basis for a reasonable explanation of these differences. There was no significant interaction found between class level and test format.

For the purpose of identifying areas of possible interest for future research, two-way ANOVAs were run for each of the items on the

Table 3 MANOVA using factor scores Wilks' multivariate test of significance

Effect	Approx F	HyP	Error	p
Format	2.58	2.00	111.00	.081
Level	2.37	6.00	306.00	.030
Level by format	1.60	6.00	222.00	.148

questionnaire. The results of these ANOVAs for those items which showed a significant difference ( $p < .05$ ) for test format, level, or the interaction of these variables are emphasized in Table 4.

There was a significant difference in student reactions to the different test formats for items 1, 4, 12, 14 and 20. An examination of means for these items revealed that students felt the pair format (2.74) provided a slightly more accurate evaluation of their ability in English than the group format (2.40). Students reported feeling somewhat more nervous during the group test (3.78) than during the pair test (3.66). Students indicated they had understood what they were required to do in the group test (4.58) more readily than in the pair test (4.27). Those receiving the group test (3.65) felt more strongly that they would prefer to be tested by the same person the next time than did those who received the pair test (3.33).

Items 3, 4, 12, 13, 14, 16 and 17 showed a significant difference among the responses of students from different levels. A review of the means showed that students from level two reported feeling more nervous before and during the tests than did students from other levels. There was not a clear pattern of differences for the remaining items, which were related either to the content of the tests, or to the test administrator. As mentioned previously, the tests were designed to assess achievement, and consequently differed in content from level to level. There were also different test administrators at each level.

Table 4 Two-way ANOVAs questionnaire items

Item	Source	Mean sq	F	p
1 Accurate evaluation of ability	Format	7.45	6.34	.013
	Level	4.23	1.70	.169
	Level by format	3.16	2.69	.049
2 Different teacher = different score	F	19	17	.683
	L	1.67	.82	.485
	LF	1.19	1.04	.374
3 Nervous before test	F	19	14	.705
	L	10.88	2.90	.037
	LF	2.10	1.59	.195
4 Nervous during test	F	6.83	5.50	.021
	L	11.47	3.72	.013
	LF	.23	.18	.909



Table 1 Crosstabulation: questionnaire item by format

Questionnaire item	Format	Disagree	No opinion	Agree
		%	%	%
1 Accurate evaluation of ability	pair	56.7	1.8	41.5
	group	66.9	0.7	32.4
2 Different teacher = different score	pair	55.2	25.8	19.0
	group	51.5	27.2	21.3
3 Nervous before test	pair	27.6	0.6	71.8
	group	32.4	1.5	66.2
4 Nervous during test	pair	27.6	0.6	71.8
	group	23.7	1.5	74.8
5 Prefer written test	pair	26.2	20.7	53.0
	group	27.2	17.6	55.1
6 Think performed well	pair	18.3	4.3	77.4
	group	16.4	2.2	81.3
7 Test is reliable	pair	34.4	8.6	57.1
	group	33.1	12.5	54.4
8 Oral tests necessary	pair	6.1	0.0	93.9
	group	3.7	0.7	95.6
9 Opportunity to show ability in English	pair	41.3	3.1	55.6
	group	44.1	2.9	52.9
10 Test too short	pair	72.2	5.6	22.2
	group	62.5	9.6	27.9
11 Liked test	pair	18.5	6.8	74.7
	group	15.4	6.6	77.9
12 Understood instructions	pair	13.7	1.9	84.5
	group	8.8	0.0	91.2
13 Test covered class material	pair	10.5	2.5	87.0
	group	3.7	4.4	91.9
14 Same teacher next time	pair	13.6	56.2	30.2
	group	8.8	49.3	41.9
15 Prefer to be tested w/ another student	pair	7.4	23.9	68.7
	group	13.3	21.5	65.2
16 Same teacher for midterm as final	pair	18.0	44.1	37.9
	group	19.1	48.5	32.4
17 Testers accent difficult to understand	pair	72.0	3.1	24.8
	group	80.0	0.7	19.3
18 Test too difficult	pair	70.0	5.0	25.0
	group	77.0	5.9	17.0
19 Like classroom teacher to be present	pair	21.6	42.0	36.4
	group	19.1	41.9	39.0
20 Prefer group to pair format	pair	38.7	24.5	36.8
	group	24.3	16.9	58.8

expressed no opinion about whether the teacher who gives the test at the midterm should be the same one who gives it at the final (item 16), although more agreed it should be the same person. Slightly over 40 per cent had no opinion about whether the classroom teacher should be present during the exam (item 19), though more students agreed he

or she should be present. In contrast, there were very few students who had no opinion about whether the tester's accent was difficult to understand (item 17), as the majority disagreed with this statement.

A few students also appeared indifferent to whether a test was written or oral. A little over half said they preferred written tests (item 5), though approximately 20 per cent expressed no opinion. At the same time, over 90 per cent said they felt oral tests were necessary (item 8).

Finally, about two thirds of the students agreed that they feel more comfortable when tested with another student (item 15), though a little over 20 per cent expressed no opinion on the matter. However, when asked specifically whether they preferred the group more than the pair format (item 20), there was a striking difference between those who had just received the pair format and those who had been tested in a group. Of those most recently tested in pairs, about an equal number (a little over a third) said they preferred the group format as those who said they did not prefer it. About a quarter of these students expressed no opinion. In contrast, over half of the students most recently tested in a group said they preferred the group format, while only a quarter said they did not prefer it.

In order to assess the effect of self-reported anxiety on performance, correlations between the total performance score and item 4 ('I felt nervous during the test') were compared. For the pair test format, a negative correlation of  $-0.27$  ( $p < .01$ ) was revealed; the correlation on the group format was also negative, at  $-0.28$  ( $p < .01$ ). Hence, there appears to have been a very similar though weak inverse relationship between self-reported anxiety and actual performance on the test for both test formats.

It was not possible to run a multivariate analysis of variance (MANOVA) using all the items as there was too much collinearity among them. Consequently, in order to reduce the data, a factor analysis (using principal factoring with varimax rotation) was performed, and the resulting factor scores were used in the MANOVA. For the factor analysis, the data from both test formats were combined into one data set, as it was assumed that the students taking either format would have interpreted questionnaire items in a similar way. In other words, students' understanding of the intent of an item was not thought to have been affected by the test format they received. Because factor plots showed items 14 and 16 to be very closely related, these items were collapsed in a subsequent factor analysis.

The factor loadings resulting from the varimax rotated factor matrix of the second factor analysis are shown in Table 2. Convergence was achieved in 10 iterations. Two factors emerged with an Eigenvalue greater than 1.00. Factor 1 had an Eigenvalue of 2.91 and Factor 2 a value of 1.30.



A teacher rating of oral proficiency (independent of the oral tests) was also obtained for each student at this time. Teachers compared each student with others at the same level and rated him or her on a five-point scale. An affective questionnaire, based on the one developed by Jones *et al.* (1980) and translated into Portuguese, was administered *after* the oral tests on each occasion. (See Appendix II for an English translation of the questionnaire. Note that all items require ratings on a five-point Likert-style scale, while some also include space for students to explain their ratings.)

### 3 Hypotheses

The primary hypothesis was that there would be a difference in students' reactions to the group and pair oral test formats. It was also hypothesized that there would be a difference in student reactions depending on their proficiency level; specifically, that students at a lower level would react more positively to the group format because of the security it was felt to provide, while students at a higher level would prefer the pair format because they would perceive it as giving them a better opportunity to demonstrate their knowledge, while still providing some security (in contrast to a one-on-one interview, where the focus would be entirely on one student). (However, as noted above, both tests provided approximately equivalent speaking opportunities to participating students.)

### 4 Analysis

Means and standard deviations were computed for all items on the affective questionnaire, and the items were cross-tabulated for format and response. Pearson Product Moment correlations were obtained for test scores and questionnaire items, and responses of students who took the pair test were compared to those of students who took the group test.

A factor analysis (using principal factoring with varimax rotation) was used to reduce the data. A multivariate analysis of variance (MANOVA), based on the factor scores obtained in the previous analysis, was then performed to assess the significance of differences between responses of students who took the group test and those who took the pair test for the items on the affect questionnaire. (Results for the midterm and final test administrations were collapsed for all of these analyses, as it was assumed that students would interpret the questionnaire items in a similar manner regardless of the administration.) In order to identify areas of possible interest for

future research, two-way ANOVAs were performed for each of the questionnaire items.

In addition, a quantitative analysis was performed on students' responses to those items on the affect questionnaire which required them to explain their ratings on the Likert-style scale. Responses to these items were translated into English and then summarized by category.

### III Results and discussion

The data on student performance at different levels were collapsed in calculating measures of central tendency. Measures of central tendency for student performance on the pair vs the group tests revealed that when grouped by format the tests did not appear to differ in their level of difficulty. The mean for performance on the pair test format ( $n = 155$ ) was 16.33 (SD = 2.88); while the mean on the group test format ( $n = 108$ ) was 16.27 (SD = 2.44). The results of Pearson Product Moment correlations revealed that for the pair test ( $n = 138$ ), the total rating of students' performance correlated with the independent teacher rating of their oral proficiency at .49 ( $p < .001$ ); for the group test ( $n = 102$ ) the correlation was .43 ( $p < .001$ ).

Students responded to an affect questionnaire following the administration of each of the test formats. Because the response scale used permitted students to express 'no opinion' (scored as a 3), it was important to determine whether an item mean close to 3 indicated that the majority of students had no opinion about the item, or whether approximately equal numbers of students disagreed as agreed with the item. Hence, Table 1 represents a cross-tabulation of item by test format. Responses for each item are summarized in three categories: disagree (responses 1 or 2), no opinion (response 3), and agree (responses 4 or 5).

As shown in Table 1, the percentage of students responding in each category (disagree, no opinion, agree) generally seems to follow similar trends for both formats, although a few differences can be observed.

Although most students generally either agreed or disagreed with an item, there was a higher percentage of students who expressed no opinion about items having to do with the test administrator. Though over half of the students didn't feel that they would have received a different score if they had taken the test from a different teacher (item 2), over a quarter of the students expressed no opinion on this issue. Half of the students expressed no opinion when asked if they would like to take the next test with the same teacher who had given it to them that day (item 14), though more students agreed they would like the same teacher than disagreed. Once again, almost half of the students



positive, even though the tests were very difficult for students. It appeared that students were able to separate their perception of difficulty from their appraisal of the fairness or validity of a test; thus, they reacted positively to a test because they felt it was valid even though they felt they had not performed well on it. Savignon concludes that although most students felt completely unprepared for the testing experience, they regarded the tests as actually testing those foreign language skills they were supposed to be learning. In a similar vein, Shohamy (1980) makes the following observations about her students' reactions to an oral interview:

Students seem to feel that the oral interview reflected their actual knowledge of the language since they could see the direct relationship between the testing procedure and their performance . . . They perceived the oral interview as a low anxiety test and some even thought that it was a fun test which made them feel comfortable and created a relaxing atmosphere. Many perceived it as a pleasant experience (pp. 15-16).

A study by Brutch (1979) also demonstrated students' concern with the validity and relevance of a testing procedure. This study examined students' attitudes toward two types of written tests: a discrete point proficiency test and a global/communicative test requiring students to write essays and letters. Students preferred the communicative test, reporting that it permitted them to show how they could apply what they had learned, and reflected real-life situations.

Research has also revealed crosscultural differences in affective reaction to tests. Scott (1980) examined the performance of Spanish and Japanese students on high and low ESL proficiency test batteries. This study showed that students from different cultural backgrounds differed in their attitude toward the test batteries as well as in their performance. There were statistically significant differences ( $p < .01$ ) between the ratings of the Spanish and Japanese students for those items which involved more emotive reactions, such as how well they liked the tests, how frustrating they considered the experience of taking the tests, their perception of their performance, and how pleasant they considered the experience of taking the tests. The Spanish students rated these items more positively than the Japanese.

The studies discussed above reveal that there are a variety of factors which influence students' affective reactions to tests. Some reactions, including the perception of test validity and level of difficulty, appear to be based on cognitive judgements, while others may reflect emotional states, such as anxiety. In addition, there seems to be variation in the performance and affective reactions of students from different cultural backgrounds to various types of tests. The aim of the present study was to examine student-affective reactions to two different oral test formats given in an

## II Method

### 1 Subjects

The data analysed in this study were collected in 1982 during the fall semester of a large EFL programme at the Associação Alumni in São Paulo, Brazil. Approximately 160 adult students participated in the study. There were students from four programme levels (levels 2, 3, 5, and 8) which represented roughly three proficiency groups: levels 2 and 3 were beginning students, level 5 were low intermediate, and level 8 were high intermediate. (No outside reference scores such as TOEFL were available for these students.) Four classes were included at each level, totaling approximately 40 students per level.

### 2 Procedure

The students were administered oral and written tests at the midterm and final. The written battery (consisting of subtests such as a multiple-choice grammar section and a cloze dialogue) were given the day before the oral tests at the midterm and the day after the oral tests for the final. For the oral tests, half of the students at each level (two classes or approximately 20 students) were tested in groups of six at the midterm and then in pairs for the final. The other half were tested in pairs at the midterm and in groups at the final (hence, each student ultimately received both the pair and the group oral test formats). Both midterm and final tests (oral and written) were level-specific achievement tests.

The group and pair oral tests were similar in content and in the type of tasks the students were asked to perform, i.e. ask and answer questions, and carry out brief role plays. In addition, students in both testing conditions had approximately the same amount of time to participate, as the groups of six students were tested for approximately 30 minutes, while the pairs were tested for about 10 minutes. Because these were achievement tests based on material covered during the course, the content varied from one level to another, although the testing formats remained consistent. In addition, each class had a different test administrator (someone other than their regular classroom teacher, usually another teacher from the same level). However, the person who actually scored the students' performance was the same for all students at each level. Students were rated on their fluency, comprehension, pronunciation, grammatical accuracy and vocabulary using a simplified scale based on the one used in the Interagency Language Roundtable test (formerly Foreign Service Institute or FSI test). (See Appendix 1 for a copy of the rating scale. The scale consisted of 100 items, which were then tallied for an overall score.



### 1 Background

Anxiety, one of the emotional reactions of students to tests, has been the object of much research. Early research on test anxiety revealed the importance of distinguishing between trait (considered a relatively stable personality characteristic) and state (more transitory, varying as a result of perceived threat) anxiety (Spielberger, 1966). According to Schwarzer, van der Ploeg, and Spielberger (1982), persons who have higher trait anxiety will tend to experience state anxiety more frequently because they perceive a wider range of situations as dangerous or threatening. Another finding has been that anxiety can have both a facilitating and/or a debilitating effect on performance (Alpert and Haber, 1960). It is interesting to note that these constructs appear to be independent of one another. As Alpert and Haber state, 'an individual may possess a large amount of both anxieties, or of one, but not the other, or of none of either' (p. 213). Although the relationships of state vs trait anxiety and the facilitating vs debilitating effects of anxiety are not entirely clear, it appears that state anxiety may have a more debilitating effect on individuals who exhibit a high degree of trait anxiety (Madsen, 1982; Hill, 1983).

A variety of factors in addition to anxiety can affect a student's performance on an exam. A study by Galassi *et al.* (1984) found little connection between test anxiety and student performance on an achievement exam. They did find, however, that a significant although small (4 per cent) amount of exam variance could be accounted for by student concerns (assessed at the beginning, middle and end of the exam) about their performance, the testing environment and time constraints. Schwarzer *et al.* (1982) have suggested that 'test anxiety may play an indirect role in performance, depending on the amount of time available for completing an examination and the degree to which an individual has prepared for the task at hand' (p. 6).

The effect of time constraints on performance is clearly illustrated in a study by Hill (1983). Hill found that under time pressure, children identified as high test anxious made three times as many errors and took twice as long as low anxious children. However, when time limits were removed, the high test anxious children performed as well as their low test anxious peers and completed the test in approximately the same amount of time. As Hill points out, it is evident that the performance of the high test anxious children in the first testing situation was limited by anxiety resulting from time constraints, and not by low achievement in the subject being tested.

A study by Madsen (1982), examining the debilitating effect of anxiety on performance on ESL examination batteries, concluded that the anxiety generated by a high anxiety-producing subtest was due

to the complexity and difficulty of the items. He commented further that 'anecdotal accounts and research indicate that in addition to the form of the exam (cloze, oral interview, etc.), faulty instructions, lack of face validity, difficulty level, insufficient time, and flagrant cheating by other students are a few of the factors that can cause anxiety and frustration while taking a test' (p. 142).

Madsen and Murray (1984) administered the Alpert-Haber *Achievement Anxiety Test* to a group of graduate students and a group of precollege ESL students. They then conducted informal interviews with the graduate students after their course midterm, and the ESL students following their placement exam. Their findings showed that high anxiety students in both groups were disturbed by similar factors, including time limitations and exam length. In addition, they found other sources of stress to be unfamiliar question types and perceived weaknesses in items or exam formats.

It appears that even the opportunity to express how one feels about a test, such as provided in the study by Madsen and Murray, may have an effect on a student's level of anxiety and performance in a testing situation. McKeachie *et al.* (1955) and Calvin *et al.* (1957) found that giving students an opportunity to write comments about objective questions resulted in improved performance on the last half of the test. McKeachie *et al.* conclude, 'Giving students an opportunity to write comments aids not only in reducing the threat but also in channeling the release of anxiety' (p. 98).

Research in language testing has also compared student affective reactions to different test formats. Jones *et al.* (1980) examined the affective reaction of students to seven different types of language tests. A factor analysis of their data revealed that the affective values assessed by their questionnaire appeared to load almost entirely on two major factors. One of these factors was identified as emotive reactions concerning how pleasant the experience of taking the tests was, how easy the tests were, how frustrating they seemed, how well the students felt they performed, and how well the students liked the tests. The other factor was identified as cognitive reactions and involved judgements as to how well the test differentiated among students, how well they reflected student proficiency in the language, how reliable the tests were, and how well they corresponded to previous classroom instruction. Student evaluation of the fairness of the tests loaded relatively highly on both factors, although somewhat more heavily on the emotive factor. Similar findings were revealed in a subsequent study (Scott, 1980) employing an expanded version of the questionnaire developed by Jones *et al.*

Savignon (1972) reports that reactions to oral tests of communicative competence examined in her study were 'overwhelmingly'



## Student affective reactions to oral language tests

Mary Lee Scott *University of California, Los Angeles*

Many factors influence student reactions to tests, including format, length, time constraints, testing environment, familiarity with test format, perceptions of test validity, and student anxiety. The present study assessed the affective reactions of native Brazilian students to different oral EFL test formats (group and pair) in an achievement testing situation. Each student participated in both the group and the pair test formats and completed an affect questionnaire after each test. A principal factor analysis of student responses to the questionnaire revealed two major factors. The first appeared to involve cognitive judgements about the validity of the test and how well students liked the test. The other factor seemed to be emotive, and involved student reports of the anxiety they felt both before and after the test. Similarly, a qualitative analysis of student responses to open-ended items on the questionnaire found that students were mostly concerned about their emotional state, test content, and time constraints in the testing situation, while comments about test administrators accounted for only a small percentage of the total responses. A multivariate analysis of variance based on the results of the factor analysis showed no significant difference among student reactions to the different test formats.

A number of recent studies have examined the affective reactions of students to various types of language tests (Savignon, 1972; Brutch, 1979; Jones *et al.*, 1980; Scott, 1980; Shohamy, 1982; Scott and Madsen, 1983; Madsen, 1982; Madsen and Murray, 1984). These studies have revealed many factors which influence student reactions to a test, including format, length, time constraints, testing environment, familiarity with test format, perceptions of test validity, difficulty and clarity of test instructions, and anxiety. In addition, researchers have discovered crosscultural differences in student performance and reactions to different test formats (Scott, 1980; Scott and Madsen, 1983). The purpose of the present study was to assess the affective reactions of native Brazilian students to different oral EFL test formats (group vs pair) in an achievement testing situation.